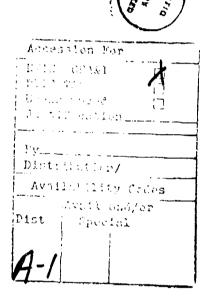Some Clinical Diagnoses Are More Reliable Than Others

Brock Kilbourne, Ph.D.

Jerry Goodman, Ph.D.

Susan Hilton, M.A.

Health Services Research Department

Naval Health Research Center

P.O. Box 85122

San Diego, CA. 92138-9174

Summary

Purpose

The purpose of the present study was to facilitate understanding of the nature of mental disorder diagnoses by examining the extent to which the degree of diagnostic specificity (i.e., group, type, or subtype) and the social context (i.e., certain circumstances under which diagnoses take place) affect diagnostic agreement (the reliability of measures across time). Also, the present study investigated the transformative nature of certain mental disorder diagnoses (i.e., how some mental disorder diagnoses change by their very nature) by tracking classifications both into and out of particular diagnostic categories.

Approach

The approach of this study was to analyze data from the Navy Enlisted Career/Medical History File. The population consisted of all hospitalized cases of active duty, enlisted Navy personnel between 1981 and 1984, inclusive, with a mental problem as the primary diagnosis. Two nonexclusive subsamples were investigated. The first subsample (N=2,132) consisted of hospitalized cases which were later reviewed by a U.S. Navy Physical Evaluation (P.E.) Board. The second subsample (N=5,402) consisted of cases involving multiple hospital admissions for a mental disorder.

Findings

Diagnostic group (i.e., psychotic versus nonpsychotic) was a more reliable measure than diagnostic type (e.g., schizophrenia versus personality disorder). Diagnostic type, in turn, was a more reliable measure than diagnostic subtype (e.g., chronic catatonic schizophrenic). And, certain diagnostic types and subtypes were consistently more reliable across time than others. Except for alcoholism and personality disorders, the Kappa values associated with diagnostic types and subtypes in the P.E. Board subsample were significantly larger than those in the multiple hospitalization subsample.

Conclusions

The present study found that degree of diagnostic specificity and the social context affected the reliability across time of ICD-9 mental disorder diagnoses. Also, the nature of certain mental disorders (i.e., disorders which progressively deteriorate, those which resemble one another because of shared features, and those which occur in combination with other disorders) affects diagnostic reliability.

Some Clinical Diagnoses Are More Reliable Than Others[1]

Brock Kilbourne, Ph.D.[2]

Jerry Goodman, Ph.D.

Susan Hilton, M.A.

Naval Health Research Center

Introduction

In a recent article in Science, Faust and Ziskin (July, 1988) argued correctly that many studies have demonstrated the poor reliability of specific mental disorder diagnoses, but then also created the false impression that: 1) all mental disorder diagnoses are similarly unreliable, and 2) the inaccuracy of diverse forms of clinical judgement, as indicated by the inability of clinicians to achieve diagnostic agreement, to predict violence, feigned behavior, or brain damage, is attributable to a common set of limitations (i.e., the same factors explain problems in very different and complex areas of clinical judgement). Spitzer, Williams, and Pincus (Science, November, 1988) responded by saying that psychiatry has recognized the problem, is doing something about it, and that the "rest of medicine also has problems with reliability." Joseph Matarazzo (quoted by John Bales, Monitor, January, 1989), the current President of the American Psychological Association, responded by saying that Faust and Ziskin's (1988) conclusions were based on a narrow review of the literature (i.e., primarily negative studies) and ignored other studies which do not suffer from methodological flaws. However, neither the critique by Faust and Ziskin (1988), the defense by Spitzer et al. (1988), nor the corrective comments by Matarazzo (Bales, 1989) provide a clear agenda for understanding the nature of mental disorder diagnoses, or suggest ways to improve diagnostic reliability.

The purpose of the present study was to facilitate understanding of the nature of mental disorder diagnoses by examining the extent to which the degree of diagnostic specificity (i.e., group, type, or subtype) and the social context (i.e., the different conditions under which diagnoses take place) affects diagnostic agreement (i.e., the reliability of measures across time). Also, the present study investigated the transformative nature of certain mental disorder diagnoses (i.e., how some diagnoses may change by their very nature) by tracking classifications both into and out of particular diagnostic categories.

Methods

Subjects

The population consisted of all hospitalized cases of active duty, enlisted Navy personnel between 1981 and 1984, inclusive, with a mental problem as the primary diagnosis. Two non-exclusive subsamples of this population were investigated, which corresponded to different social contexts. The first social context subsample (N=2,132) consisted of hospitalized cases which were later reviewed by a U.S. Navy Physical Evaluation (P.E.) Board. The mean length of time between hospitalization and last P.E. Board review was 206.74 days (sd=214.45). The second social context subsample (N=5,402) consisted of cases involving multiple hospital admissions for a mental disorder. The mean lengths of time between hospitalizations were: 1) 143.55 days (sd=225.08) between first and second hospitalization, 2) 115.82 days (sd=187.59) between second and third hospitalization, 3) 91.00 days (sd=149.42) between third and fourth hospitalization, 4) 91.45 days (sd=139.05) between fourth and fifth hospitalization, and 5) 176.78 days (sd=251.13) between first and last hospitalization.

Procedures

Data Collection Procedures. Data were obtained from the Navy Enlisted Career/Medical History File (NECMHF). NECMHF is based on two compiled files. One is the Service History File, which consists of demographic and military-service history data from Navy Military Personnel Command in Arlington, Virginia. The other is the Medical History File, which contains hospitalization, death, Medical Board action, and Physical Evaluation Board action data from Naval Medical Data Services Cen r in Bethesda, Maryland. NECMHF is compiled and maintained by the Naval Health Research Center, San Diego, California (Garland, Helmkamp, Gunderson, Gorham, Miller, McNally, & Thompson, 1987).

Coding Procedures. Primary mental disorders were based on ICD-9 codes and were coded to correspond to three levels of diagnostic specificity. The three levels, which are henceforth referred to as **group, type,** and **subtype,** correspond, respectively, to low, moderate, and high degrees of diagnostic specificity. At the lowest level of diagnostic specificity, the **group** level, all mental disorder diagnoses were relegated to either the psychotic (all psychotic diagnoses combined) or nonpsychotic (all nonpsychotic diagnoses combined) groups. At the moderate level of specificity, the **type** level, the range of possible mental disorder diagnoses included the following diagnostic types: 1) organic psychoses, 2) schizophrenia, 3) affective psychoses, 4) para-

noia, 5) other psychoses, 6) unspecified psychoses, 7) neurotic disorders, 8) personality disorders, 9) alcoholism, 10) transient situational disturbance, and 11) drunkenness. At the highest level of specificity, the **subtype** level, each **type** of mental disorder diagnosis was further classified into its specific **subtypes** (e.g., chronic catatonic schizophrenic). In both subsamples, five contrasts involving the three levels of diagnostic specificity (**group, type,** and **subtype**) were computed to assess diagnostic reliability across time: 1) psychotic group versus nonpsychotic group at Time 1 was compared to psychotic group versus nonpsychotic group at Time 2 (2 x 2 Table), 2) all types of mental disorders at Time 1 were compared to all types of mental disorders at Time 2, 3) all subtypes of mental disorders at Time 1 were compared to all subtypes of mental disorders at Time 2, 4) each type of mental disorder versus all other types combined at Time 1 was compared, respectively, to its corresponding type of mental disorder versus all other types combined at Time 2 (2 x 2 Tables), and 5) each subtype of mental disorder versus all other subtypes combined at Time 1 was compared, respectively, to its corresponding subtype of mental disorder versus all other subtypes combined at Time 2 (2 x 2 Tables).

Results

The degree to which diagnosis at Time 1 agreed with diagnosis at Time 2 was assessed using the Kappa statistic. Kappa assesses the chance-corrected consistency (or reliability) of a set of measurements across time. Kappa equals zero when obtained agreement equals chance agreement. Greater than chance agreement leads to positive Kappa values, while less than chance agreement leads to negative Kappa values. With negative values, the degree of agreement has little practical importance (Cohen, 1960). Table 1 shows the Kappa values for different levels of diagnostic specificity across multiple hospital admissions and at time of hospitalization versus time of last P.E. Board review.

Two sets of analyses were conducted. The first set of analyses combined the P.E. Board subsample with the multiple hospital admissions subsample in order to assess the effect of diagnostic specificity on reliability across time. The Wilcoxon Signed-Ranks Test was used to compare the Kappa values for different levels of diagnostic specificity. The Kappa values comparing psychotic group versus nonpsychotic group at Times 1 and 2 were significantly

larger than the Kappa values comparing: 1) all types of diagnoses at Time 1 versus all types of diagnoses at Time 2, and 2) all subtypes of diagnoses at Time 1 versus all subtypes of diagnoses at Time 2 (both Z values=-2.20, p values[two-tailed]<.05). The Kappa values comparing all types of diagnoses at Time 1 versus all types of diagnoses at Time 2 were significantly larger than those comparing all subtypes of diagnoses at Time 1 versus all subtypes of diagnoses at Time 2 (Z=-2.20, p[two-tailed]<.05). The Kappa values (Tables 2 and 3) for each respective diagnostic type versus all other types at Times 1 and 2 were significantly larger than those for each respective diagnostic subtype versus all other subtypes at Times 1 and 2 (Z=-5.03, p[two-tailed]<.001).

The second set of analyses combined the diagnostic types and subtypes in the P.E. Board subsample, and combined the diagnostic types and subtypes in the multiple hospital admissions subsample. The Wilcoxon Signed-Ranks Test was used to compare the combined Kappa values of the two different subsamples in order to assess the effect of the social context on reliability across time. In subsample 1, Kappa values were computed to measure reliability at time of hospitalization versus time of last P.E. Board review. In subsample 2, Kappa values were computed to measure reliability at time of first hospital admission versus time of last hospital admission. The Kappa values of the two subsamples (types and subtypes combined, respectively) did not differ significantly (p=.16). However, two supplemental analyses were conducted to assess the possibility that opposing directional effects in the two subsamples masked a social context effect. The first supplemental analysis consisted of only the types and subtypes with larger P.E. Board subsample Kappa values (12 of the 16 ranks or all types and subtypes except alcoholism and personality disorders) and found a significant difference between the two subsamples (Z=-3.06, p[two-tailed] <.01). The second supplemental analysis consisted of only alcoholism and personality disorder types and subtypes (the remaining 4 ranks) and found a significant difference between the two subsamples (Z=-1.83, p[one-tailed] <.05). Except for alcoholism and personality disorders, Kappa values were significantly larger in the P.E. Board subsample than in the multiple hospitalization subsample. The Kappa values for alcoholism and personality disorder types and subtypes were significantly larger in the multiple hospitalization subsample than in the P.E. Board subsample. Thus, the impact of the social context on diagnostic reliability depended on the particular diagnostic type or subtype.

Table 1

Reliability (Kappa) Across Time of Mental Diagnosis by Level of Specificity

Level of specificity	Number of Hospital Admissions					Row Mean	Time of Hospitalization vs. Last PE Board
	1 vs 2	2 vs 3	3 vs 4	4 vs 5	1st vs Last		
Group vs. Group	.56	.57	.53	.67	.54	.57	.55
Type vs. Type	.36	.37	.43	.44	.34	.39	.52
Subtype vs. Subtype	.01	.14	.02	.03	.13	.07	.01

Table 2

Reliability (Kappa) Across Time of Mental Diagnosis Type

Type vs. Else	Number of Hospital Admissions					Time of Hospitalization vs. Last PE Board
	1 vs 2	2 vs 3	3 vs 4	4 vs 5	1st vs Last	
Organic Psychoses	.15	.06	.16	.15	.13	.31
Schizophrenia	.58	.61	.63	.73	.58	.72
Affective Psychoses	.42	.57	.54	.59	.36	.61
Other Psychoses	.25	.30	.16	.66	.21	.35
Neuroses	.26	.12	.11	-.03	.22	.54
Personality Disorders	.35	.36	.41	.40	.32	.16
Alcoholism	.52	.51	.47	.58	.50	.14
Transient Situational Disturbance	.24	.22	.39	.26	.20	.33
Drunkenness	.10	.10	.49	.26	.09	--

Table 3

Reliability (Kappa) Across Time of Mental Diagnosis Subtype

Subtype vs. Else		Number of Hospital Admissions					Time of Hospitalization vs. Last PE Board
		1 vs 2	2 vs 3	3 vs 4	4 vs 5	1st vs Last	
Organic Psychoses	(14 subtypes)	.14	.07	.07	.09	.10	.23
Schizophrenia	(11 subtypes)	.44	.46	.47	.54	.44	.55
Affective Psychoses	(9 subtypes)	.29	.36	.29	.41	.26	.40
Other Psychoses	(13 subtypes)	.19	.23	.17	.66	.16	.32
Neuroses	(8 subtypes)	.23	.13	.13	-.02	.19	.50
Personality Disorders	(9 subtypes)	.29	.37	.44	.40	.26	.15
Alcoholism	(6 subtypes)	.42	.42	.43	.50	.39	.08
Transient Situational Disturbance	(12 subtypes)	.10	.10	.49	.26	.09	.21
Drunkenness	(1 subtype)	.17	.16	.32	.12	.15	--

The pattern of diagnostic change was then assessed by examining classifications across time both into and out of a given type of mental disorder. The pattern of diagnostic change of primary mental disorder diagnoses was different in the two subsamples of the study. Tables 4-5 show the shift in diagnosis from time of hospitalization to time of last P.E. Board review. Tables 6-7 show the shift in diagnosis from time of first hospital admission to time of last hospital admission.

In the P.E. Board subsample, diagnostic change appeared to be due to the problem of differential diagnosis of psychotic mental disorders (the failure to distinguish between different mental disorders which share particular psychotic symptoms but which differ in the overall pattern of psychotic symptomatology) and the progressive deterioration of some nonpsychotic mental disorders. For example, Table 4 clearly shows that at time of last P.E. Board review, four of the six psychotic mental disorders (#1,4,5,6) were not well differentiated from schizophrenia and that all of the nonpsychotic disorders (#7-11) indicated some deterioration into a more severe mental (psychotic) disorder. Table 5 clearly shows a consistent trend toward classification into a new category of diagnosis at time of last P.E. Board review but cannot distinguish between problems of differential diagnosis and progressive deterioration.

In the multiple admission subsample, on the other hand, diagnostic change was more complex and appeared to be due to: 1) the problem of differential diagnosis for both psychotic and nonpsychotic disorders, 2) the problem of multiple diagnosis (clusters of disorders [e.g., certain personality disorders and substance abuse disorders] which are oftentimes related and which may alternate as the principal [primary] diagnosis), 3) and the progressive deterioration of some nonpsychotic disorders. For example, Table 6 shows clearly that schizophrenic and/or personality disorders were frequently confused with all seven psychotic mental disorders (#1-7). Table 6 also shows substantial diagnostic change among the nonpsychotic disorders that could be attributable to progressive deterioration, differential diagnosis, or multiple diagnosis. The pattern of diagnostic change for drunkenness (#11), improper use of drugs (#13), and nervousness/debility (#17) looks like movement toward a worsening of symptoms. The pattern of change for neuroses (#8), drug dependence (#12), physical disorder of psychogenic origin (#14), special symptoms (#15), and transient situational disturbance (#16) could be attributable to

8

Table 4

**Classification Out of a Diagnostic Type
at Time of Last P.E. Board**

Hospitalization Diagnosis	P.E. Board Diagnosis	
(1) Psychosis associated with other physical conditions (n=24)	Psychosis associated with other physical conditions	17%
	Schizophrenia	46%
	Affective psychosis	25%
(2) Schizophrenia (n=777)	Schizophrenia	94%
	Affective psychosis	3%
(3) Affective psychosis (n=352)	Affective psychosis	84%
	Schizophrenia	6%
	Nervousness/debility	4%
(4) Paranoid (n=46)	Paranoid	41%
	Schizophrenia	26%
(5) Other psychosis (n=67)	Other psychosis	25%
	Schizophrenia	43%
(6) Unspecified psychosis (n=93)	Unspecified psychosis	26%
	Schizophrenia	54%
(7) Neuroses (n=150)	Neuroses	62%
	Schizophrenia	10%
	Affective psychosis	8%
(8) Personality disorders (n=160)	Personality disorders	13%
	Schizophrenia	32%
	Affective psychosis	18%
	Neuroses	11%
(9) Alcoholism (n=90)	Alcoholism	8%
	Schizophrenia	20%
	Affective psychosis	31%
(10) Transient situational disturbance (n=135)	Transient situational disturbance	25%
	Affective psychosis	21%
	Schizophrenia	14%
	Neuroses	19%
(11) Nervousness/debility (n=129)	Nervousness/debility	27%
	Affective psychosis	30%
	Neuroses	11%

Table 5

**Classification Into a Diagnostic Type
at Time of Last P.E. Board**

	Hospitalization Diagnosis		P.E. Board Diagnosis
(1)	Schizophrenia	74%	Schizophrenia (n=984)
	Unspecified psychosis	5%	
	Personality disorders	5%	
	Other psychosis	3%	
	Affective psychosis	3%	
(2)	Affective psychosis	59%	Affective psychosis (n=502)
	Transient situational disturbance	6%	
	Personality disorders	6%	
	Alcoholism	6%	
(3)	Paranoid	46%	Paranoid (n=41)
	Personality disorders	13%	
(4)	Other psychosis	33%	Other psychosis (n=52)
	Personality disorders	12%	
	Schizophrenia	8%	
(5)	Unspecified psychosis	43%	Unspecified psychosis (n=56)
	Other psychosis	11%	
	Schizophrenia	11%	
(6)	Neuroses	54%	Neuroses (n=172)
	Transient situational disturbance	15%	
	Nervousness/debility	11%	
	Personality disorders	1%	
(7)	Personality disorders	39%	Personality disorders (n=52)
	Affective psychosis	15%	
	Paranoid	13%	
	Schizophrenia	12%	
(8)	Physical disorder of psychogenic origin	57%	Physical disorder of psychogenic origin (n=23)
	Neuroses	30%	
(9)	Special symptoms, not elsewhere classified	78%	Special symptoms, not elsewhere classified (n=41)
	Alcoholism	10%	
	Personality disorders	7%	
(10)	Transient situational disturbance	60%	Transient situational disturbance (n=57)
	Nervousness/debility	18%	
(11)	Nervousness/debility	37%	Nervousness/debility (n=95)
	Transient situational disturbance	16%	
	Affective psychosis	15%	
	Neuroses	13%	

Table 6

**Classification Out of a Diagnostic Type
at Time of Last Hospitalization**

First Hospitalization Diagnosis	Last Hospitalization Diagnosis	
(1) Alcoholic psychosis (n=65)	Alcoholism	74%
	Schizophrenia	9%
(2) Psychosis associated with other physical conditions (n=37)	Psychosis associated with other physical conditions	14%
	Schizophrenia	32%
	Alcoholism	16%
(3) Schizophrenia (n=247)	Schizophrenia	72%
	Personality disorders	12%
(4) Affective psychosis (n=122)	Affective psychosis	43%
	Personality disorders	19%
(5) Paranoid (n=21)	Paranoid	14%
	Schizophrenia	33%
	Personality disorders	29%
(6) Other psychosis (n=74)	Other psychosis	14%
	Schizophrenia	26%
	Personality disorders	22%
(7) Unspecified psychosis (n=78)	Unspecified psychosis	9%
	Schizophrenia	11%
	Personality disorders	22%
	Affective psychosis	10%
(8) Neuroses (n=273)	Neuroses	22%
	Personality disorders	29%
	Alcoholism	15%
	Transient situational disturb.	13%
(9) Personality disorders (n=545)	Personality disorders	57%
	Alcoholism	13%
	Transient situational disturb.	7%
	Schizophrenia	6%
(10) Alcoholism (n=1,517)	Alcoholism	72%
	Personality disorders	8%
	Drunkenness	5%
	Transient situational dist rb.	5%
(11) Drunkenness (n=311)	Drunkenness	11%
	Alcoholism	71%
	Personality disorders	8%
(12) Drug dependence (n=54)	Drug dependence	26%
	Alcoholism	26%
	Personality disorders	13%
(13) Improper use of drugs (n=29)	Improper use of drugs	14%
	Alcoholism	31%
	Personality disorders	17%
(14) Physical disorder of psychogenic origin (n=27)	Physical disorder of psychogenic origin	30%
	Personality disorders	26%
	Alcoholism	22%
(15) Special symptoms, not else where classified (NEC) (n=28)	Special symptoms, NEC	54%
	Personality disorders	25%
(16) Transient situational distur bance (n=525)	Transient situational disturb.	23%
	Personality disorders	35%
	Alcoholism	18%
	Neuroses	7%
(17) Nervousness/debility (n=114)	Nervousness/debility	20%
	Personality disorders	26%
	Transient situational disturb.	15%
	Alcoholism	13%

11

Table 7

Classification Into a Diagnostic Type
at Time of Last Hospitalization

First Hospitalization Diagnosis		Last Hospitalization Diagnosis
(1) Alcoholic psychosis	13%	Alcoholic psychosis (n=46)
Alcoholism	67%	
(2) Psychosis associated with other physical conditions	21%	Psychosis associated with other physical conditions (n=24)
Alcoholism	25%	
Drug dependence	17%	
(3) Schizophrenia	53%	Schizophrenia (n=336)
Personality disorders	10%	
Unspecified psychosis	9%	
Other psychosis	6%	
(4) Affective psychosis	37%	Affective psychosis (n=144)
Transient situational disturb.	11%	
Personality disorders	10%	
Nervousness/debility	10%	
(5) Other psychosis	22%	Other psychosis (n=46)
Personality disorders	28%	
Alcoholism	13%	
(6) Unspecified psychosis	20%	Unspecified psychosis (n=35)
Alcoholism	14%	
Other psychosis	11%	
Schizophrenia	11%	
(7) Neuroses	32%	Neuroses (n=185)
Transient situational disturb.	21%	
Alcoholism	19%	
Personality disorders	11%	
(8) Personality disorders	35%	Personality disorders (n=895)
Transient situational disturb.	21%	
Alcoholism	14%	
Neuroses	9%	
(9) Alcoholism	67%	Alcoholism (n=1,643)
Drunkenness	13%	
Transient situational disturb.	6%	
Personality disorders	4%	
(10) Drunkenness	21%	Drunkenness (n=157)
Alcoholism	50%	
Personality disorders	9%	
Transient situational disturb.	7%	
(11) Drug dependence	30%	Drug dependence (n=47)
Alcoholism	26%	
(12) Improper use of drugs	10%	Improper use of drugs (n=40)
Alcoholism	30%	
Drug dependence	18%	
(13) Physical disorder of psychogenic origin	38%	Physical disorder of psychogenic origin (n=21)
Neuroses	33%	
(14) Special symptoms, not elsewhere classified (NEC)	65%	Special symptoms, NEC (n=23)
Neuroses	13%	
Personality disorders	13%	
(15) Transient situational disturb.	37%	Transient situational disturbance (n=323)
Alcoholism	22%	
Personality disorders	11%	
Neuroses	11%	
(16) Nervousness/debility	26%	Nervousness/debility (n=91)
Transient situational disturb.	20%	
Personality disorders	12%	
Neuroses	12%	

either differential diagnosis or multiple diagnoses. The consistent associa-
tion of personality disorders (#9) and alcoholism (#10) suggests multiple
diagnoses.

Discussion

The present study found that degree of diagnostic specificity and the
social context affected the reliability across time of ICD-9 mental disorder
diagnoses. Diagnostic group (i.e., psychotic versus nonpsychotic) was a more
reliable measure than diagnostic type (e.g., schizophrenic disorder or person-
ality disorder). Diagnostic type, in turn, was a more reliable measure than
diagnostic subtype (e.g., chronic catatonic schizophrenic). And, certain
diagnostic types and subtypes (e.g., schizophrenic and chronic catatonic
schizophrenic, respectively) were consistently more reliable across time than
others (e.g., personality disorder and antisocial personality disorder,
respectively). Except for alcoholism and personality disorders, the Kappa
values associated with diagnostic types and subtypes in the P.E. Board subsam-
ple (time of hospitalization versus time of last P.E. Board review) were sig-
nificantly larger than those in the multiple hospitalization subsample (i.e.,
time of first hospital admission versus time of last hospital admission).
That social context effect was attributable to the fact that Navy P.E. Boards
screen for more severe disorders as a primary basis for service discharge
(Kilbourne, Hilton, and Goodman, 1988). Moderate to large Kappa values in the
P.E. Board subsample indicated relative stability in the severity of a given
mental disorder (especially for psychotic types and subtypes), while lower
Kappa values generally indicated a shift from a less severe mental disorder to
a more severe mental disorder. For example, alcoholism is treated on either
an outpatient or inpatient basis in the U.S. Navy (e.g., there are treatment
programs lasting for six weeks in many Navy hospitals, which increases the
likelihood of diagnostic agreement while in the hospital) and is not routinely
used as the basis for service discharge (which decreases the likelihood of
diagnostic agreement if an individual initially hospitalized for alcoholism is
later reviewed by a Navy P.E. Board).

Thus, the nature of certain mental disorders (e.g., disorders which pro-
gressively deteriorate, those which resemble one another because of shared
features, and those which occur in combination with other disorders) affects
diagnostic reliability. Diagnostic change and uncertainty (i.e., symptom
change and symptom overlap) are an inherent aspect of some mental disorders.

13

It may only be after a battery of psychological tests has been administered or treatment has begun that a final diagnosis can be specified with some degree of confidence. Even then, the clinician might have to concede that the final diagnosis is not the only possible diagnosis but the most plausible working diagnosis. Notwithstanding, while problems with differential and multiple diagnosis are two areas that clinicians need to work on to improve diagnostic reliability (two areas that are emphasized in the revised third-edition of the Diagnostic and Statistical Manual of Mental Disorders [DSM-III-R, 1987]), it is also evident that diagnostic change can be affected by certain organizational settings and policies.

Taken together, these findings suggest that in actual practice a "close enough principle" steers most clinical diagnoses, and is probably the general rule regardless of the classification system or range of diagnostic types examined. That observation is by no means alarming since effective treatment is not contingent upon diagnostic exactness. The matching of specific psychotherapies or specific medications to specific mental disorders is not critical for the successful treatment of most mental disorders. Treatment success has resulted, in large part, from the nonspecific effects of therapy (e.g., expectations of help and hope, new learning experiences, successful outcomes, and increased self-mastery [Frank, 1974; Sloan, Stables, Cristol, Yorkston, & Whipple, 1975; Smith & Glass, 1977; Stiles, Shapiro, & Elliot, 1986]) and/or the general effects associated with a particular class of medication (e.g., neuroleptics, antidepressants, or sedatives/hypnotics). An individual suffering from a particular mental disorder may respond positively to any one of many different forms of psychotherapeutic intervention (e.g., psychodynamic, cognitive-behavioral, humanistic, etc.).

When specificity of treatment (e.g., behavioral or medical) has been indicated by the weight of empirical evidence, such specificity is really only approximate in application. For example, there are several techniques of the behavioral approach or different forms of a particular class of medication which produce similar effects, and the same technique or form of medication is not always administered in the same way. More importantly, such approximate specificity in application does not hold true for all people at all times with a particular mental disorder.

Thus, for all practical purposes, the belief that one has hit the diagnostic "bull's eye" may impose limitations upon the full range of psychotherapeutic intervention. On the other hand, the recognition that mental diagnoses

14

are inexact, but close enough, assures appropriate boundaries around the problem (e.g., the focus of treatment) as well as sufficient flexibility, given cultural and individual differences, in treating the problem. The acceptable range of diagnostic closeness is an empirical question that remains to be determined.

The therapeutic alliance (the special relationship that develops between the therapist and the client [Frank, 1974; Strupp, 1986; Kilbourne & Richardson, 1988]) is more important in the long run for helping most individuals deal with their psychological problems than either diagnostic or treatment exactness. It is within the context of the therapeutic alliance and certain common clinical strategies (e.g., corrective experiences and feedback [Goldfried, 1980]) that an individual learns new ways to tackle unmet needs and unresolved conflicts, distorted perceptions, irrational beliefs, negative affective states, faulty communication patterns, inappropriate behaviors, unrewarding relationships, and to acquire a positive sense of self.

Footnotes

[1] An earlier version of this paper was presented in a poster session at the Annual Meeting of the American Association for the Advancement of Science, San Francisco, CA., January 17, 1980.

[2] Brock Kilbourne is a postdoctoral fellow with the National Research Council, National Academy of Sciences, and a licensed psychologist (CA. #PV10467). Jerry Goodman is a sociologist and a statistical consultant with the Naval Health Research Center, San Diego. Susan Hilton is a research psychologist and a member of the Health Services Research Department, Naval Health Research Center.

References

Bales, J. (January, 1989). Expert witness controversy continues. APA Monitor. 20(1), 21.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20, 37-46.

Diagnostic and statistical manual of mental disorders (Third Edition Revised) (1987). Washington, DC.: American Psychiatric Association.

Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. Science, 241, 31-35.

Frank, J. (1974). Persuasion and healing: A comparative study of psychotherapy. New York: Schocken.

Garland, F.C., Helmkamp, J.C., Gunderson, E.K.E., Gorham, E.D., Miller, M.R., McNally, M.S., & Thompson, F.A. (1987). A guide to the computerized medical data resources of the Naval Health Research Center (NHRC Tech. Rep. 87-13). San Diego, CA: Naval Health Research Center.

Goldfried, M.R. (1980). Toward the delineation of therapeutic change principles. American Psychologist, 35, 991-999.

Kilbourne, B.K., Hilton, S.M., & Goodman, J. (1988). Psychiatric and non psychiatric predictors of disability discharge disposition for Navy personnel with a mental health problem: A replication and extension. (NHRC Tech. Rep. 88-40). San Diego, CA: Naval Health Research Center.

Kilbourne, B.K., & Richardson, J.T. (1988). A social psychological analysis of healing. Journal of Integrative and Eclectic Psychotherapy, 1, 20-34.

Sloan, R.B., Stables, F.R., Cristol, A.H., Yorkston, N.J., & Whipple, K. (1975). Psychotherapy versus behavior therapy. Cambridge, MA: Harvard University Press.

Smith, M.L., & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32, 752-760.

Spitzer, R.L., Williams, J.B., & Pincus, H.A. (4 November, 1988). Psychiatric Diagnosis. Science, 242, 651-652.

Stiles, W.B., Shapiro, D.A., & Elliot, R. (1986). Are all psychotherapies equivalent? American Psychologist, 41(2), 165-180.

Strupp, H.H. (1986). Psychotherapy: Research, practice, and public policy (how to avoid dead ends). American Psychologist, 41(2), 120-130.

17

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified			1b RESTRICTIVE MARKINGS None		
2a SECURITY CLASSIFICATION AUTHORITY N/A			3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited		
2b DECLASSIFICATION/DOWNGRADING SCHEDULE N/A					
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NHRC Report No. 89 - 11			5 MONITORING ORGANIZATION REPORT NUMBER(S)		
6a NAME OF PERFORMING ORGANIZATION Naval Health Research Center	6b OFFICE SYMBOL *(If applicable)* 40		7a NAME OF MONITORING ORGANIZATION Commander, Naval Medical Command		
6c ADDRESS *(City, State, and ZIP Code)* P.O. Box 85122 San Diego, CA 92138-9174			7b ADDRESS *(City, State, and ZIP Code)* Department of the Navy Washington, DC 20372		
8a NAME OF FUNDING/SPONSORING ORGANIZATION Naval Medical Research & Development Command	8b OFFICE SYMBOL *(If applicable)*		9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Naval Military Personnel Command Reimbursable Doc #N0002288WRWW508, Appro 1781804		
8c ADDRESS *(City, State, and ZIP Code)* Naval Medical Command National Capital Region Bethesda, MD 29814-5044			10 SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO.

11 TITLE *(Include Security Classification)*

(U) SOME CLINICAL DIAGNOSES ARE MORE RELIABLE THAN OTHERS

12 PERSONAL AUTHOR(S)
Brock Kilbourne, Jerry Goodman, and Susan Hilton

13a TYPE OF REPORT Interim	13b TIME COVERED FROM _____ TO _____	14. DATE OF REPORT *(Year, Month, Day)* 1989 March 29	15 PAGE COUNT

16. SUPPLEMENTARY NOTATION

17	COSATI CODES		18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)* Mental Disorder Diagnoses Reliability Clinical Judgement
FIELD	GROUP	SUB-GROUP	

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The purpose of the present study was to facilitate understanding of the nature of mental disorder diagnoses by examining the extent to which the degree of diagnostic specificity and the social context impact on diagnostic agreement. Also, the present study investigated the transformative nature of certain mental disorder diagnoses. The population consisted of all hospitalized cases of active duty, enlisted Navy personnel between 1981 and 1984, inclusive, with a mental problem as the primary diagnosis. Two nonexclusive subsamples were investigated. Diagnostic group was a more reliable measure than diagnostic type. Diagnostic type, in turn, was a more reliable measure than diagnostic subtype. And, certain diagnostic types and subtypes were consistently more reliable across time than others. Kappa values associated with diagnostic types and subtypes differed significantly in the two samples. Thus, degree of diagnostic specificity and the social context affected the reliability across time of ICD-9 mental disorder diagnoses.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a NAME OF RESPONSIBLE INDIVIDUAL Brock Kilbourne	22b TELEPHONE *(Include Area Code)* (619) 553-9968	22c OFFICE SYMBOL 40

DD FORM 1473, 84 MAR · 83 APR edition may be used until exhausted · SECURITY CLASSIFICATION OF THIS PAGE

All other editions are obsolete